Journal of Cardiology 67 (2016) 133–139

Contents lists available at ScienceDirect

Journal of Cardiology

journal homepage: www.elsevier.com/locate/jjcc

Original article

Whole exome sequencing combined with integrated variant annotation prediction identifies a causative myosin essential light chain variant in hypertrophic cardiomyopathy

Akihiro Nomura (MD)[a,b,c,d,1], Hayato Tada (MD, PhD)[a,1], Ryota Teramoto (MD)[a,1], Tetsuo Konno (MD, PhD, FJCC)[a,e,*], Akihiko Hodatsu (MD, PhD)[a], Hong-Hee Won (PhD)[b,c,d,f], Sekar Kathiresan (MD)[b,c,d,f], Hidekazu Ino (MD, PhD, FJCC)[a], Noboru Fujino (MD, PhD, FJCC)[a], Masakazu Yamagishi (MD, PhD, FJCC)[a], Kenshi Hayashi (MD, PhD, FJCC)[a]

[a] Division of Cardiovascular Medicine, Kanazawa University Graduate School of Medical Science, Kanazawa, Japan
[b] Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA
[c] Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA
[d] Department of Medicine, Harvard Medical School, Boston, MA, USA
[e] Research and Education Center for Innovative and Preventive Medicine, Kanazawa University, Kanazawa, Japan
[f] Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA

ARTICLE INFO

Article history:
Received 29 July 2015
Received in revised form 19 August 2015
Accepted 4 September 2015
Available online 9 October 2015

Keywords:
Hypertrophic cardiomyopathy
Myosin essential light chain
Whole exome sequencing
Bioinformatics

ABSTRACT

Background: The development of candidate gene approaches to enable molecular diagnosis of hypertrophic cardiomyopathy (HCM) has required extensive and prolonged efforts. Whole exome sequencing (WES) technologies have already accelerated genetic studies of Mendelian disorders, yielding approximately 30% diagnostic success. As a result, there is great interest in extending the use of WES to any of Mendelian diseases. This study investigated the potential of WES for molecular diagnosis of HCM.

Methods: WES was performed on seven relatives from a large HCM family with a clear HCM phenotype (five clinically affected and two unaffected) in the Kanazawa University Hypertrophic Cardiomyopathy Registry. Serial bioinformatics filtering methods as well as using combined annotation dependent depletion (CADD) score and high heart expression (HHE) gene data were applied to detect the causative variant. Moreover, additional carriers of the variant were investigated in the HCM registry, and clinical characteristics harboring the variant were collected and evaluated.

Results: WES detected 60020 rare variants in the large HCM family. Of those, 3439 were missense, nonsense, splice-site, or frameshift variants. After genotype–phenotype matching, 13 putative variants remained. Using CADD score and HHE gene data, the number of candidates was reduced to one, a variant in the myosin essential light chain (MYL3, NM_000258.2:c.281G>A, p.Arg94His) that was shared by the five affected subjects. Additional screening of the HCM registry (n = 600) identified two more subjects with this variant. Serial assessments of the variant carriers revealed the following phenotypic characteristics: (1) disease-penetrance of 88%; (2) all clinically affected carriers exhibited asymmetric septal hypertrophy with a substantial maximum left ventricular wall thickness of 18 ± 3 mm without any obstruction.

Conclusions: WES combined with CADD score and HHE gene data may be useful even in HCM. Furthermore, the MYL3 Arg94His variant was associated with high disease penetrance and substantial interventricular septal hypertrophy.

DOI of commentary article: http://dx.doi.org/10.1016/j.jjcc.2015.10.002

* Corresponding author at: Research and Education Center for Innovative and Preventive Medicine, Kanazawa University, 13-1 Takaramachi, Kanazawa 920-8641, Japan. Tel.: +81 76 265 2251; fax: +81 76 234 4251.
E-mail address: konnokontetsu@staff.kanazawa-u.ac.jp (T. Konno).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jjcc.2015.09.003
0914-5087/

Introduction

Hypertrophic cardiomyopathy (HCM) is a genetically heterogeneous myocardial disorder with various morphological, functional, and clinical characteristics [1–5]. Familial HCM, which has a prevalence of up to 1 in 500 individuals, is one of the most common autosomal dominant inherited disorders [6]. Myosin heavy chain (MYH7, OMIM#: 160760), myosin binding protein C (MYBPC3, OMIM#: 600958), troponin T (TNNT2, OMIM#: 191045), tropomyosin (TPM1, OMIM#: 191010), and troponin I (TNNI3, OMIM#: 191044) have been reported to be the main causal genes [2–4,7–11]. To date, candidate-gene approaches using traditional Sanger sequencing have enabled the identification of causative variants in approximately 50% of HCM patients [8]. However, because these techniques are onerous and time-consuming, motivating the researchers to perform target re-sequencing with next generation sequencing (NGS) as well as whole exome sequencing (WES) which is reported to have a success rate of nearly 30% in identifying disease-causing variants in patients whose causal variants have not been detected [12]. Although several other reports have demonstrated the efficacy of target re-sequencing for sporadic HCM patients [13,14], it remains unclear whether WES enables comprehensive identification of causative variants in an exome-wide manner. This study investigated the effectiveness of WES with bioinformatics for molecular diagnosis for HCM.

Materials and methods

Study population

All subjects were enrolled through the Kanazawa University Hypertrophic Cardiomyopathy Registry in Kanazawa, Japan. The largest HCM family in the registry, whose causative variant had not been detected by traditional direct sequencing, was first investigated in this study. HCM was diagnosed according to the 2011 version of the guideline of the American College of Cardiology Foundation/American Heart Association [1]. In brief, subjects with a maximal left ventricular wall $\geq$13 mm (involving asymmetric septal hypertrophy) without extra-cardiac or metabolic findings based on echocardiography were diagnosed with HCM. Individuals were classified as HCM-affected in the case diagnosed with HCM, or HCM-unaffected in the case of either a non-HCM phenotype or known diseases such as hypertensive heart disease. Detailed clinical data, including family history, age, and symptoms at evaluation, presence of hypertension, physical examination, New York Heart Association (NYHA) classification, electrocardiogram (ECG) and echocardiographic findings were collected for each subject. Hypertrophic obstructive cardiomyopathy (HOCM) was defined as HCM with left ventricular outflow tract (LVOT) obstruction and a peak instantaneous LVOT pressure gradient of $>$30 mmHg. Maximum left ventricular wall thickness (MWT) was defined as the greatest thickness within the chamber. Other echocardiographic parameters were evaluated using the recommendation of the American Society of Echocardiography [15]. Hypertension was defined as when a patient's systolic blood pressure (BP) was $\geq$140 mmHg and/or its diastolic BP was $\geq$90 mmHg. Pathological Q waves were defined as follows based on previous studies: Q wave $>$1/4 of the ensuing R wave in depth and/or $>$40 ms in duration in at least two leads except aVR [16]. T waves $>$10 mm in depth in any leads were defined as giant negative T waves (GNTW) [17]. The Ethics Committee for Medical Research at our institution approved the study protocol, and all subjects provided written informed consent.

Exome sequencing

Genomic DNA was isolated from peripheral white blood cells of all subjects using a standard DNA extraction protocol. DNA was pooled, selected size, ligated to sequencing adapters, and amplified to enrich for targets to be sequenced by the Agilent SureSelect^{XT} Target Enrichment System (Agilent Technologies Inc., Santa Clara, CA, USA). Exome capture was performed with the Agilent SureSelect^{XT} Human All Exon 50 Mb Kit (Agilent Technologies Inc.). Exome enriched products were sequenced using the Illumina HiSeq 2000 by Takara Bio Inc., Shiga, Japan. One sample was sequenced per lane to obtain an average theoretical depth of 80x, using 2x100 bp sequencing.

Bioinformatics

For the samples, paired-end reads were aligned using the Burrows-Wheeler Aligner on the human reference genome build hg19 using quality score calibration, soft clipping, and adapter trimming. Following the exclusion of PCR duplicate reads using the Picard, insertions/deletions and single-nucleotide polymorphisms (SNP) were called using the Genome Analysis Toolkit (GATK) [18,19]. Variants (SNP/indels) were filtered on the basis of the Phred scaled genotype quality score. Re-alignment was performed and the calling algorithm merged the output of GATK Unified Genotyper. All samples were annotated using SnpEff [20] (version 3.6) to classify variants (*e.g.* missense, stop gain/loss, splice-site variant, synonymous, intronic, insertions/deletions).

In addition to the standard variant quality controls, six independent filters were applied to facilitate detection of causal variants among the enrolled HCM families. Variants were filtered by: (1) minor allele frequency (MAF) $>$1% in Asian population; (2) benign, as predicted by SnpEff; (3) genotype–phenotype unmatched under the assumption of complete penetrance without phenocopies; (4) registered in the SNP Database (dbSNP137); (5) combined annotation dependent depletion (CADD) score $<$10; and (6) low heart expression of the genes, less than the top quartile.

The frequency filter adopted the allele frequency estimates from the Asian cohort of the 1000 Genomes Project database [21], and we used a MAF 1% as the cut-off. Genotype–phenotype matching was defined as narrowing the variants at which affected subjects had $\geq$1 alternative allele(s) and unaffected subjects had no alternative allele. As a functional filtering method, variants not registered in NCBI dbSNP137 were considered to be candidate variants. Prediction of *in silico* pathogenicity for novel missense variants was performed using the CADD prediction software (version 1.0), which objectively integrates many diverse annotations into a single measure (C-score) for each variant [22]. A variant was predicted to be pathogenic if the scaled C-score calculated by the software was above 10, a score indicative of the variant being within the top 10% of deleteriousness substitutions. Candidate variants were evaluated if the gene associated with each variant was directly involved in the myocardium or worsening cardiac function by using the high heart expression (HHE) gene data as previously reported [23]. In brief, HHE genes were defined as the top quartile [$>$40 reads per million mapped reads (rpm)] of 16,599 human-mouse orthologous gene expressions generated by RNA sequencing data of mice hearts at embryonic day 14.5.

After the evaluation, Sanger sequencing method was performed to confirm the putative variant identified by the bioinformatics analysis in the tested subjects and other relatives.

Additional screening for HCM registry

In addition, restriction fragment length polymorphism (RFLP) was performed on HCM probands listed in our registry to identify other subjects harboring the variant. In brief, restriction enzyme (*Van91I*) was added to each DNA sample and incubated for 3 h, which allowed the enzyme to cut at the recognition site (in this case, the myosin essential light chain [(MYL3, OMIM#: 160790),

Table 1
Clinical features of carriers harboring the MYL3 (Arg94His) variant in HCM-F18 and HCM-F189 families.

Pedigree ID	Gender	Age[a]	HT	LVOT obstruction	Symptoms	NYHA	ECG findings	QW	GNTW	LAD	MWT	IVS	PW	LVDd	LVDs	LVEF
HCM-F18																
III:3	Male	74	–	–	–	I	1°AV-block, CLBBB	–	–	47	19	19	8	55	33	61
III:6 (proband)	Female	60	–	–	–	I	AF, LVH	–	–	61	20	20	10	38	25	64
III:9	Female	58	–	–	–	I	LVH	–	–	28	23	23	11	38	23	71
III:12	Male	51	–	–	Syncope	I	1°AV-block, NSVT	–	–	53	18	18	11	48	33	59
IV:6	Female	37	–	–	–	I	–	–	–	18	10	10	9	38	23	65
IV:11	Female	24	–	–	–	I	IRBBB	–	–	30	15	15	11	39	21	72
HCM-F189																
II:2	Male	75	–	–	–	I	LVH	–	+	35	18	18	12	43	26	65
III:5	Male	41	–	–	–	I	LVH	–	–	28	13	13	10	43	30	52

AF, atrial fibrillation; ASH, asymmetric septal hypertrophy; CLBBB, complete left bundle branch block; ECG, echocardiography; GNTW, giant negative T wave; HT, hypertension; IRBBB, incomplete right bundle branch block; IVS, interventricular septum diameter; LAD, left atrial dimension; LVDd, left ventricular diastolic diameter; LVDs, left ventricular systolic diameter; LVEF, left ventricular ejection fraction; LVH, left ventricular hypertrophy; LVOT, left ventricular outflow tract; MWT, max wall thickness; NSVT, non-sustained ventricular tachycardia; NYHA, New York Heart Association heart failure classification; PW, posterior wall diameter; QW, Q wave.
[a] Age at examination.

NM_000258.2:c.281G>A variant site]). The DNA fragments were separated by size using gel electrophoresis to allow confirmation of cutting of the sample DNAs. To investigate the ancestral origin of the variant, haplotype analysis was performed by using six microsatellite markers (short tandem repeats [$(CA)_n$: D3S3687, D3S3678, D3S3647, D3S3582, D3S3640, D3S1568] flanking the variant gene) [24,25]. Clinical characteristics of variant carriers were documented for phenotype evaluation of the variant.

Results

Characteristics of HCM-F18 family members

Clinical characteristics of HCM-F18 family members are shown in Table 1, and a representative case (III:6, proband) is presented in Fig. 1. The subject was referred to our hospital because of an abnormal ECG finding detected by a local hospital. ECG showed atrial fibrillation with left ventricular hypertrophy pattern, chest X-rays displayed an increase in cardiothoracic ratio, and echocardiography revealed an apparent interventricular septal hypertrophy without LVOT obstruction. After excluding other cardiac diseases, the subject was diagnosed with HCM.

Although the HCM-F18 proband and her family members had been clinically observed for years, traditional Sanger sequencing failed to identify the apparent causative variants. The HCM-affected family members had a typical clinical phenotype consistent with HCM, exhibiting interventricular septal hypertrophy and asymmetric septal hypertrophy. The affected subjects comprised two males and three females with an age range at enrollment of between 24 and 74 years. The HCM-unaffected subjects were a 67-year-old female diagnosed with hypertensive heart disease and a 48-year-old female without any cardiac disease.

Exome sequencing and bioinformatics analyses

DNA samples from seven members of the HCM-F18 family, five clinically HCM-affected subjects, and two unaffected subjects were analyzed using WES (Fig. 2) followed by bioinformatics filtering methods (Table 2). The mean sequencing depth for the seven subjects was 56.1x, 60.0x, 54.8x, 58.7x, 58.2x, 63.7x, and 66.0x per base across the whole exome, for samples III:2 (unaffected), III:3 (affected), III:6 (affected), III:9 (affected), III:12 (affected), IV:7 (unaffected), and IV:11 (affected), respectively. Percentages of on-target reads were 78.6% (III:2), 78.0% (III:3), 77.3% (III:6), 79.4% (III:9), 77.2% (III:12), 78.7% (IV:7), and 77.0% (IV:11), respectively, while coverage rates of target coding lesions (20x) were 89.3% (III:2), 90.1% (III:3), 89.0% (III:6), 90.8% (III:9), 89.7% (III:12), 91.7% (IV:7), and 92.0% (IV:11), respectively.

Subsequent to WES, bioinformatics analysis and genotype–phenotype matching following exome sequencing were performed for family HCM-F18 to identify causative variants found in all five HCM-affected subjects but not in the two unaffected subjects (Table 2). The number of aligned variants in the seven subjects with

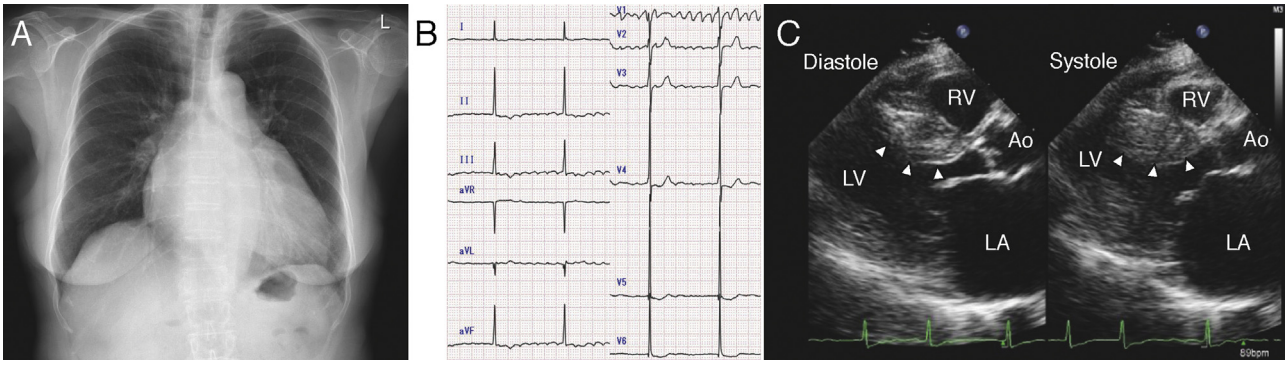

Fig. 1. A representative case (proband) in hypertrophic cardiomyopathy (HCM)-F18 family. Clinical images of the proband (III:6 in Table 1) in HCM-F18 family are shown: (A) chest X-ray; (B) electrocardiogram; (C) echocardiography of longitudinal views. Chest X-ray showed that the cardiothoracic ratio was increased (61%); electrocardiogram showed atrial fibrillation with left ventricular hypertrophy pattern; and echocardiography presented normal ejection fraction and an apparent interventricular septal hypertrophy without left ventricular outflow tract obstruction (arrow head). Ao, aorta; LA, left atrium; LV, left ventricle; RV, right ventricle.

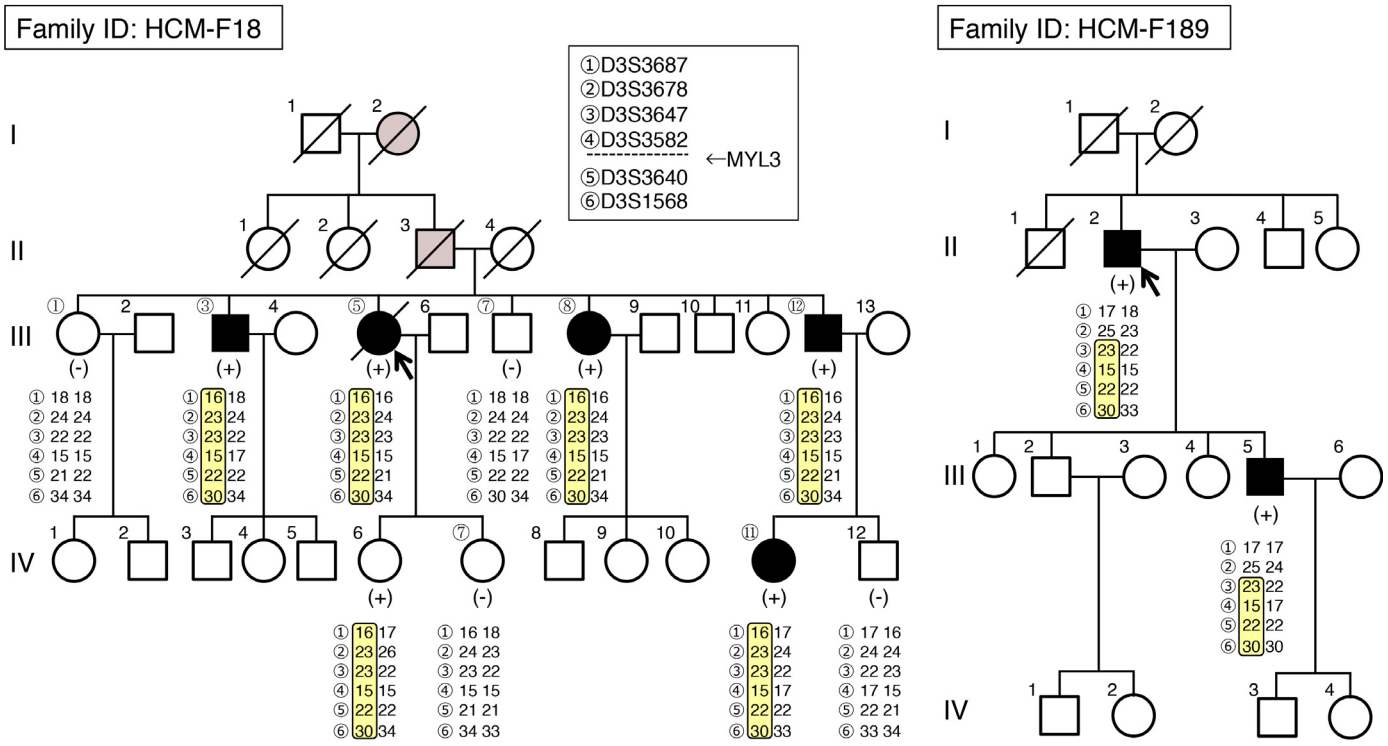

Fig. 2. Pedigrees of families: (A) hypertrophic cardiomyopathy (HCM)-F18 and (B) HCM-F189. The pedigrees of HCM families with the myosin essential light chain (MYL3) variant (Arg94His) are shown. Arrows indicate probands. Squares represent male subjects; circles represent female subjects. Diagonal lines mark deceased individuals. Solid symbols denote HCM. Open symbols represent unaffected individuals or individuals with no data available for analysis. Gray symbols represent obligate carriers. Plus "(+)" and minus "(−)" signs indicate presence or absence of the Arg94His variant confirmed by direct sequencing, respectively. Encircled numbers indicate subjects investigated by whole exome sequencing. Haplotypes (the number of $[CA]_n$ repeat) are listed for six microsatellite markers ([1] D3S3687 – [2] D3S3678 – [3] D3S3647 – [4] D3S3582 – [MYL3] – [5] D3S3640 – [6] D3S1568) from top to bottom. Boxed haplotypes demonstrate affected haplotypes. Affected haplotypes share same numbers of $[CA]_n$ repeat from [3] D3S3647 to [6] D3S1568 in both families, indicating that the MYL3 Arg94His variant originated from the same ancestor.

passing quality filtering was 243359. Among those, 60200 were considered to be rare variants (MAF <1%) using the Asian cohort in 1000 Genome Project [21]. Focusing on missense, nonsense, splice-site, and frameshift variants, 3439 variants were detected. Genotype–phenotype matching enabled putative variants to narrow the number to 13. After excluding variants registered in dbSNP137, six putative variants remained (Table 3). To evaluate these variants, CADD prediction software was used to calculate each scaled C-score. Five out of six variants were predicted to be damaging (scaled C-score >10), indicating that these were potentially deleterious variants. Among the candidate variants, myosin essential light chain (MYL3, c.281G>A, p.Arg94His) was the only gene that belonged to the HHE genes and was directly associated with the myocardium (ventricular muscle). Although this MYL3 Arg94His variant was not listed in the NHLBI ESP exome variant server [26], the Human Genetic Variation Database [27], or the Exome Aggregation Consortium database [28], the variant was reported to be a possibly damaging variant in the Human Gene Mutation Database [29]. Sanger sequencing for the 10 HCM-F18 family members validated the variant in all the affected subjects and resulted in the detection of one additional variant carrier without an apparent HCM phenotype (Figs. 2 and 3).

Table 2
Algorithm of bioinformatics filtering methods.

Total aligned variants in seven subjects	257452
After QC	243359
	↓
Filtering methods	
MAF <1% in the 1000 Genome Project (Asian cohort)	60020
	↓
Missense, nonsense, splice-site or frameshift variants	3439
	↓
Genotype–phenotype matching	13
	↓
Remove dbSNP137 registered variants	6
	↓
CADD score >10	5
	↓
High heart expression gene	1

The number of variants was decreased according to each filtering step. Finally, 1 variant, MYL3 c.281G>A was thought to be the causative variant.
CADD, combined annotation dependent depletion; MAF, minor allele frequency; QC, quality control; dbSNP, the Single Nucleotide Polymorphism Database.

Phenotype evaluation of MYL3 (c.281G>A, p.Arg94His) variant

All 600 familial or sporadic HCM probands listed in the registry were screened for the MYL3 (c.281G>A, p.Arg94His) variant by RFLP analysis using the *Van91I* restriction enzyme. The Arg94His variant was identified in one additional proband (HCM-F189) and Sanger sequencing confirmed that two additional subjects had the MYL3 Arg94His variant (Fig. 2). Moreover, haplotype analysis of family HCM-F18 and family HCM-F189 using six microsatellites in the neighborhood revealed that the MYL3 Arg94His variant originated from the same ancestor (Fig. 2).

The clinical characteristics of the eight MYL3 Arg94His variant carriers are summarized in Table 1. Disease-penetrance was 88%, and all clinically affected carriers exhibited asymmetrical septal hypertrophy with a substantial maximal left ventricular wall thickness of 18 ± 3 mm. Left ventricular systolic dysfunction (ejection fraction <50%) was not observed. Abnormal ECG findings were observed in seven carriers with an apparent HCM phenotype.

Table 3
Candidate variants after filtering methods in HCM-F18.

Gene	Function	Chr.	Exon	Position (build 37)	NM #	Amino acid	SIFT	PolyPhen-2	MutationTaster2	CADD score	HHE gene
HMGB4	High Mobility Group Box 4	1	2	34329932	145205	E47A	0.01	0.999	DC	24.5	–
HHATL	Hedgehog Acyltransferase-Like	3	9	42738359	20707	R341C	0	1	DC	18.7	–
MYL3	**Myosin Light Chain 3, Ventricular, Skeletal**	**3**	**3**	**46902192**	**258**	**R94H**	**0.04**	**0.008**	**DC**	**16.63**	+
NIPAL4	NIPA-Like Domain Containing 4	5	1	156887258	1172292	R39Q	0.25	0.002	Poly	19.5	–
PLAU	Plasminogen Activator, Urinary	10	10	75676220	1145031	R381L	0.11	0.009	Poly	9.587	–
LIPM	Lipase, Family member M	10	4	90574372	1128215	G184S	0.01	0.981	DC	35	–

The six variants in six genes filtered by bioinformatics methods are present in five HCM-affected subjects and are absent in two unaffected subjects. Of those, 5 variants had the CADD score more than 10, and MYL3 gene was the only gene that highly expressed in mice hearts.
SIFT scores (ranges from 0 to 1) were calculated by SIFT version 5.2.2, and the SIFT score less than 0.05 was considered as deleterious. PolyPhen-2 scores (ranges from 0 to 1) were calculated by PolyPhen-2 version 2.2.2, and the PolyPhen-2 score more than 0.85 was considered as probably damaging. The six variants were also evaluated by MutationTaster2 (classifying variants as "Disease causing" or "Polymorphism").
CADD, combined annotation dependent depletion; Chr, chromosome; DC, disease causing; HCM, hypertrophic cardiomyopathy; HHE, high heart expression; Poly, polymorphism; PolyPhen, Polymorphism Phenotyping v2 score; Position, nucleotide position; SIFT, Sorting Intolerant From Tolerant score.
Bold highlights the causative variant.

Three subjects had left ventricular hypertrophy, two had first-degree atrioventricular block, two had bundle branch block, and one had atrial fibrillation with an enlarged left atrium. GNTW was observed in one subject, but pathological Q wave was not observed.

Clinical follow-up data were available for four of eight subjects. Although they suffered from neither cardiac death nor progression of left ventricular systolic dysfunction during >10 years follow-up, one subject (HCM-F18; III:12) received an implantable cardioverter defibrillator (ICD) because of syncope with frequent non-sustained ventricular tachycardia. Another subject (HCM-F18; III:6) died from non-cardiac causes (postoperative infection after an orthopedic surgery).

Discussion

In this study, WES was demonstrated to be an effective tool, even in HCM. Furthermore, we determined the MYL3 Arg94His variant was newly identified as a variant associated with high disease penetrance and substantial degree of interventricular septal hypertrophy for the first time. This is the first study to validate the MYL3 Arg94His as a causative variant of familial HCM using WES.

Since the application of NGS methods such as WES and target re-sequencing for clinical genetics, the number of bioinformatics approaches for detecting the causative variant in patients in

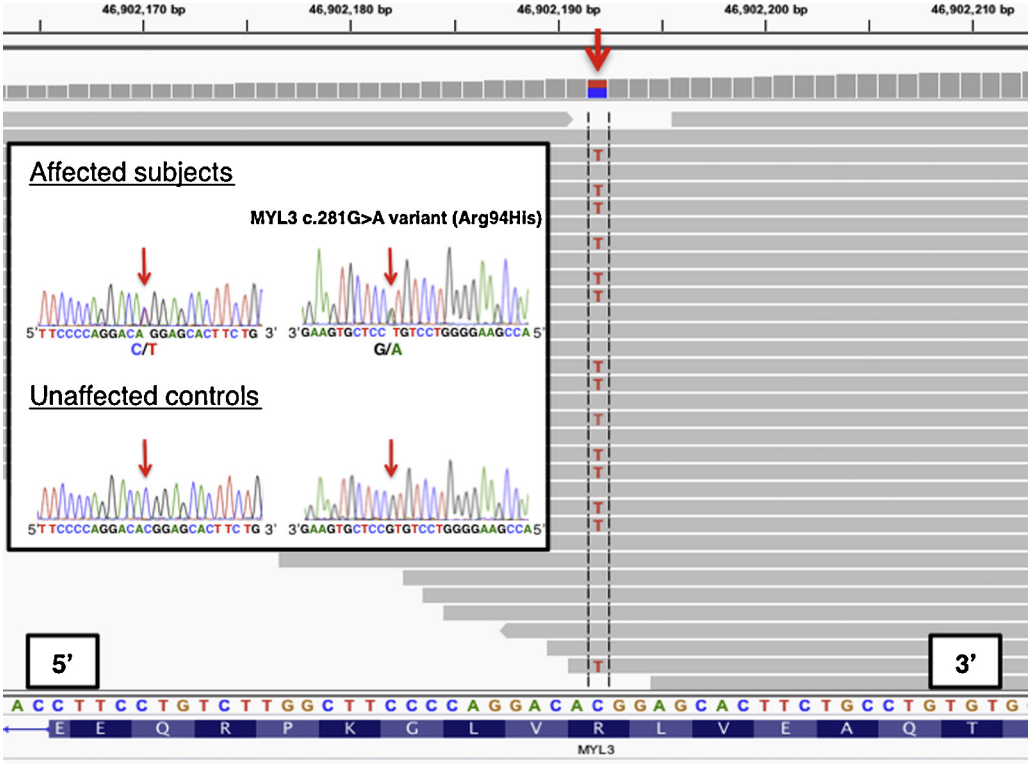

Fig. 3. NGS read alignment and the Sanger sequencing validation in the MYL3 gene at exon 3. Alignment reads are visualized by the Integrative Genomics Viewer (IGV) version 2.3. The red arrow in IGV indicates the p.Arg94His variant position in the MYL3 gene at exon 3 (5′ → 3′ complementary strand). The red arrow in Sanger sequencing data of affected subjects also shows the same variant at the same position.

cardiomyopathy has gradually increased [30,31]. This is because the traditional direct sequencing methods have only been able to identify approximately 50% of the causative variants in patients with cardiomyopathy [8]. Several clinical studies have demonstrated that WES is effective for molecular diagnosis of cardiomyopathy [30,31], providing support for the notion that WES may be useful in diverse ethnic settings.

In this study, two novel filtering schemas were adopted to enable causative variants to be distinguished from thousands of other incidental variants detected *via* WES. The CADD C-score [22] is a highly combined damaging score for each variant that is generated by 63 distinct annotation tools including SIFT [32] and PolyPhen [33]. This means that we can obtain the comprehensive damaging scores of each candidate variant even though each annotation tool shows discordant results. For example, the MYL3 Arg94His variant was first predicted as deleterious by SIFT and disease-causing by MutationTaster2 [34], but was considered as benign by PolyPhen-2 (Table 3). However, the CADD C-score of the variant was 16.63, clearly predicting that the MYL3 Arg94His as deleterious. Furthermore, prediction of the pathogenicity of human variants with the CADD C-score is more accurate than that of any other *in silico* single annotation tool [22]. Actually, the usefulness of CADD C-score for detecting causative variants in patients with Tangier disease and long QT syndrome were previously reported [35,36]. In terms of HHE [23], Zaidi et al. demonstrated that HHE genes had a higher frequency of protein-alternating *de novo* variants than genes with low heart expression (less than top quartile) in patients with congenital heart diseases [23]. In the present study, the combination of the CADD scaled C-score and the heart expression-oriented filtering schema successfully narrowed down the number of causative variants to a single variant, the MYL3 Arg94His variant, suggesting the usefulness of this schema even in heterogeneous diseases such as HCM.

Completion of the final filtering method (Table 2) indicated that MYL3 was the only gene that was directly associated with ventricular myocardium. MYL3 encodes the essential light chain, an important component of sarcomere, which wraps around the lever arm of the myosin head, and supports both the neck domain and the lever arm with myosin regulatory light chain [37]. Previous reports demonstrate that the loss of MYL3 function causes familial HCM [38–43]. Olson et al. described HCM patients with the autosomal recessive MYL3 Glu143Lys variant [38]. These patients exhibited mid-cavity hypertrophy with restrictive physiology, but in none of our HCM-subjects was the feature found. Also, Andersen et al. and Kazmierczak et al. reported the MYL3 Val79Ile and Ala57Gly variants were causative in HCM [41,43]. In this study, all subjects with HCM harbored the MYL3 Arg94His variant. Although the Arg94His was suggested as a potentially causative variant of sporadic HCM by DNA re-sequencing array [44] or Sanger sequencing [45], and was also classified as a "pathogenic" variant according to the American College of Medical Genetics and Genomics guideline [46], this is the first study to confirm and support it as a causative variant of HCM by using the co-segregation pattern in the HCM pedigrees and the unbiased method of WES. In addition, a detailed description of the clinical phenotype associated with the MYL3 Arg94His is provided. The myosin light chain belongs to the EF-hand family of calcium-binding proteins [47]. Therefore, loss-of EF-hand function may interfere with binding of calcium or magnesium causing malfunction of myosin kinetics [37], which is thought to be a potential pathological mechanism of HCM development [42]. The MYL3 causative variants are primarily located in exons 3 and 4, which encode the EF-hand 2 domain [37], and the Arg94 residue is also located in exon 3. Use of the CADD scaled C-score [22] also predicted this to be a damaging variant site. The MYL3 Arg94His variant is expected to cause crucial damage to EF-hand function, myosin kinetics, and its structure, and eventually hypertrophic changes in the myocardium, which may underlie the high disease penetrance and substantial interventricular septal hypertrophy observed in this study (Table 1). However, due to the small number of subjects harboring this variant, it remains unclear whether screening of families with HCM for the MYL3 Arg94His variant will be useful in patient management. Further study is needed to clarify the long-term clinical courses in HCM patients with the MYL3 Arg94His variant.

Study limitations

This study has some limitations. WES was performed with protein-coding regions only, which potentially overlooked other intergenic variants that influenced the specific phenotype. HCM is a heterogeneous disorder in terms of both clinical characteristics and age of onset, limiting the accuracy of the phenotype modeling. In this study, we could accurately find out one causative variant for HCM, but it might be controversial whether a 48-year-old woman without apparent HCM phenotype could be classified as an unaffected HCM subject. Although CADD enabled more precise prediction of *in silico* damaging score than other software, the optimal cut-off value for detection of deleterious variants with CADD scaled C-scores is debatable. HHE genes were generated only by known human-mouse orthologues, which might rule-out potential human-specific genes that were highly expressed in human heart. In addition, clinical information was not available for some family members, which may have affected the penetrance and clinical characteristics of HCM with MYL3 Arg94His variant.

Conclusion

Together with CADD score and HHE gene data, WES facilitated successful identification of the one causative variant, the MYL3 Arg94His, even in HCM. Furthermore, the MYL3 Arg94His variant was associated with high disease penetrance and substantial ventricular hypertrophy.

Funding

This study was supported by JSPS KAKENHI Grant Numbers 19590807, 22590808, 25460648, and 26460742. Noboru Fujino was funded by a grant for innovative research from Kanazawa University Hospital. Akihiro Nomura was also funded by a research grant from Yoshida Scholarship Foundation.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors express their gratitude to Drs Christine E. Seidman, Jonathan G. Seidman, and Jason Homsy for providing heart expression RNA-sequencing data from mice. Drs Hidenobu Terai, Tomoya Kaneda, Yoshihiro Tanaka, Yoji Nagata, Toyonobu Tsuda, and Mika Mori for assistance in collecting follow-up information including clinical and echocardiographic data, and Takako Obayashi for technical assistance in collecting genetic information and laboratory experiments.

References

[1] Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, Seidman CE, Towbin JA, Udelson JE, Yancy CW, American College of Cardiology Foundation/American Heart Association Task Force on Practice G. 2011 ACCF/AHA Guideline for the Diagnosis and Treatment of Hypertrophic Cardiomyopathy: a report of the American

College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Developed in collaboration with the American Association for Thoracic Surgery, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2011;58:e212–60.
[2] Maron BJ, Maron MS. Hypertrophic cardiomyopathy. Lancet 2013;381:242–55.
[3] Fujino N, Konno T, Hayashi K, Hodatsu A, Fujita T, Tsuda T, Nagata Y, Kawashiri MA, Ino H, Yamagishi M. Impact of systolic dysfunction in genotyped hypertrophic cardiomyopathy. Clin Cardiol 2013;36:160–5.
[4] Kawashiri MA, Hayashi K, Konno T, Fujino N, Ino H, Yamagishi M. Current perspectives in genetic cardiovascular disorders: from basic to clinical aspects. Heart Vessels 2014;29:129–41.
[5] Fujita T, Konno T, Yokawa J, Masuta E, Nagata Y, Fujino N, Funada A, Hodatsu A, Kawashiri MA, Yamagishi M, Hayashi K. Increased extent of myocardial fibrosis in genotyped hypertrophic cardiomyopathy with ventricular tachyarrhythmias. J Cardiol 2015;66:63–8.
[6] Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. Circulation 1995;92:785–9.
[7] Kimura A, Harada H, Park JE, Nishi H, Satoh M, Takahashi M, Hiroi S, Sasaoka T, Ohbuchi N, Nakamura T, Koyanagi T, Hwang TH, Choo JA, Chung KS, Hasegawa A, et al. Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. Nat Genet 1997;16:379–82.
[8] Maron BJ, Maron MS, Semsarian C. Genetics of hypertrophic cardiomyopathy after 20 years: clinical perspectives. J Am Coll Cardiol 2012;60:705–15.
[9] Lopes LR, Rahman MS, Elliott PM. A systematic review and meta-analysis of genotype–phenotype associations in patients with hypertrophic cardiomyopathy caused by sarcomeric protein mutations. Heart 2013;99:1800–11.
[10] Kokado H, Shimizu M, Yoshio H, Ino H, Okeie K, Emoto Y, Matsuyama T, Yamaguchi M, Yasuda T, Fujino N, Ito H, Mabuchi H. Clinical features of hypertrophic cardiomyopathy caused by a Lys183 deletion mutation in the cardiac troponin I gene. Circulation 2000;102:663–9.
[11] Terauchi Y, Kubo T, Baba Y, Hirota T, Tanioka K, Yamasaki N, Furuno T, Kitaoka H. Gender differences in the clinical features of hypertrophic cardiomyopathy caused by cardiac myosin-binding protein C gene mutations. J Cardiol 2015;65:423–8.
[12] Stitziel NO, Peloso GM, Abifadel M, Cefalu AB, Fouchier S, Motazacker MM, Tada H, Larach DB, Awan Z, Haller JF, Pullinger CR, Varret M, Rabes JP, Noto D, Tarugi P, et al. Exome sequencing in suspected monogenic dyslipidemias. Circ Cardiovasc Genet 2015;8:343–50.
[13] Lopes LR, Zekavati A, Syrris P, Hubank M, Giambartolomei C, Dalageorgou C, Jenkins S, McKenna W, Uk10k C, Plagnol V, Elliott PM. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. J Med Genet 2013;50:228–39.
[14] Gomez J, Reguero JR, Moris C, Martin M, Alvarez V, Alonso B, Iglesias S, Coto E. Mutation analysis of the main hypertrophic cardiomyopathy genes using multiplex amplification and semiconductor next-generation sequencing. Circ J 2014;78:2963–71.
[15] Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, Picard MH, Roman MJ, Seward J, Shanewise JS, Solomon SD, Spencer KT, Sutton MS, Stewart WJ, Chamber Quantification Writing G. et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr 2005;18:1440–63.
[16] Konno T, Shimizu M, Ino H, Yamaguchi M, Terai H, Uchiyama K, Oe K, Mabuchi T, Kaneda T, Mabuchi H. Diagnostic value of abnormal Q waves for identification of preclinical carriers of hypertrophic cardiomyopathy based on a molecular genetic diagnosis. Eur Heart J 2004;25:246–51.
[17] Konno T, Fujino N, Hayashi K, Uchiyama K, Masuta E, Katoh H, Sakamoto Y, Tsubokawa T, Ino H, Yamagishi M. Differences in the diagnostic value of various criteria of negative T waves for hypertrophic cardiomyopathy based on a molecular genetic diagnosis. Clin Sci 2007;112:577–82.
[18] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
[19] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.
[20] Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 2012;6:80–92.
[21] 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
[22] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
[23] Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, Carriero NJ, Cheung YH, Deanfield J, DePalma S, Fakhro KA, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature 2013;498:220–3.
[24] Sakata K, Shimizu M, Ino H, Yamaguchi M, Terai H, Fujino N, Hayashi K, Kaneda T, Inoue M, Oda Y, Fujita T, Kaku B, Kanaya H, Mabuchi H. High incidence of sudden cardiac death with conduction disturbances and atrial cardiomyopathy caused by a nonsense mutation in the STA gene. Circulation 2005;111:3352–8.
[25] Thierfelder L, Watkins H, MacRae C, Lamas R, McKenna W, Vosberg HP, Seidman JG, Seidman CE. Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. Cell 1994;77:701–12.
[26] NHLBI Exome Sequencing Project (ESP). Exome Variant Server, http://evs.gs.washington.edu/EVS/.
[27] Japanese genetic variation consortium. A reference database of genetic variations in Japanese population, http://www.genome.med.kyoto-u.ac.jp/SnpDB.
[28] The Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org.
[29] Human Gene Mutation Database (HGMD), http://www.hgmd.cf.ac.uk.
[30] Campbell N, Sinagra G, Jones KL, Slavov D, Gowan K, Merlo M, Carniel E, Fain PR, Aragona P, Di Lenarda A, Mestroni L, Taylor MR. Whole exome sequencing identifies a troponin T mutation hot spot in familial dilated cardiomyopathy. PLOS ONE 2013;8:e78104.
[31] Guo Q, Xu Y, Wang X, Guo Y, Xu R, Sun K, Chen S. Exome sequencing identifies a novel MYH7 p.G407C mutation responsible for familial hypertrophic cardiomyopathy. DNA Cell Biol 2014;33:699–704.
[32] Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4.
[33] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
[34] Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014;11:361–2.
[35] Tada H, Kawashiri MA, Nohara A, Saito R, Tanaka Y, Nomura A, Konno T, Sakata K, Fujino N, Takamura T, Inazu A, Mabuchi H, Yamagishi M, Hayashi K. Whole exome sequencing combined with integrated variant annotation prediction identifies asymptomatic Tangier disease with compound heterozygous mutations in ABCA1 gene. Atherosclerosis 2015;240:324–9.
[36] Hayashi K, Konno T, Tada H, Tani S, Liu L, Fujino N, Nohara A, Hodatsu A, Tsuda T, Tanaka Y, Kawashiri MA, Ino H, Makita N, Yamagishi M. Functional characterization of rare variants implicated in susceptibility to lone atrial fibrillation. Circ Arrhythm Electrophysiol 2015 [Epub ahead of print].
[37] Moore JR, Leinwand L, Warshaw DM. Understanding cardiomyopathy phenotypes based on the functional impact of mutations in the myosin motor. Circ Res 2012;111:375–85.
[38] Olson TM, Karst ML, Whitby FG, Driscoll DJ. Myosin light chain mutation causes autosomal recessive cardiomyopathy with mid-cavitary hypertrophy and restrictive physiology. Circulation 2002;105:2337–40.
[39] Sanbe A, Nelson D, Gulick J, Setser E, Osinska H, Wang X, Hewett TE, Klevitsky R, Hayes E, Warshaw DM, Robbins J. In vivo analysis of an essential myosin light chain mutation linked to familial hypertrophic cardiomyopathy. Circ Res 2000;87:296–302.
[40] Lee W, Hwang TH, Kimura A, Park SW, Satoh M, Nishi H, Harada H, Toyama J, Park JE. Different expressivity of a ventricular essential myosin light chain gene Ala57Gly mutation in familial hypertrophic cardiomyopathy. Am Heart J 2001;141:184–9.
[41] Andersen PS, Hedley PL, Page SP, Syrris P, Moolman-Smook JC, McKenna WJ, Elliott PM, Christiansen M. A novel Myosin essential light chain mutation causes hypertrophic cardiomyopathy with late onset and low expressivity. Biochem Res Int 2012;2012:685108.
[42] Lossie J, Ushakov DS, Ferenczi MA, Werner S, Keller S, Haase H, Morano I. Mutations of ventricular essential myosin light chain disturb myosin binding and sarcomeric sorting. Cardiovasc Res 2012;93:390–6.
[43] Kazmierczak K, Paulino EC, Huang W, Muthu P, Liang J, Yuan CC, Rojas AI, Hare JM, Szczesna-Cordary D. Discrete effects of A57G-myosin essential light chain mutation associated with familial hypertrophic cardiomyopathy. Am J Physiol Heart Circ Physiol 2013;305:H575–89.
[44] Fokstuen S, Lyle R, Munoz A, Gehrig C, Lerch R, Perrot A, Osterziel KJ, Geier C, Beghetti M, Mach F, Sztajzel J, Sigwart U, Antonarakis SE, Blouin JL. A DNA resequencing array for pathogenic mutation detection in hypertrophic cardiomyopathy. Hum Mutat 2008;29:879–85.
[45] Zou Y, Wang J, Liu X, Wang Y, Chen Y, Sun K, Gao S, Zhang C, Wang Z, Zhang Y, Feng X, Song Y, Wu Y, Zhang H, Jia L, et al. Multiple gene mutations, not the type of mutation, are the modifier of left ventricle hypertrophy in patients with hypertrophic cardiomyopathy. Mol Biol Rep 2013;40:3969–76.
[46] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–23.
[47] Collins JH. Homology of myosin DTNB light chain with alkali light chains, troponin C and parvalbumin. Nature 1976;259:699–700.